



TITLE:

# 楔形文字の世界と文字認識の実現

AUTHOR(S):

山本, 孟; 森, 若葉; 山内, 健二

---

CITATION:

山本, 孟 ...[et al]. 楔形文字の世界と文字認識の実現. 京都大学アカデミックデイ2017: 研究者と立ち話 (ポスター/展示) 2017: 44.

ISSUE DATE:

2017-09-30

URL:

<http://hdl.handle.net/2433/227860>

RIGHT:



# 楔形文字の世界と文字認識の実現



表① 楔形文字の発展 (注: 音価の右下の数字は同音異綴の文字を区別するためにつけられる)

	紀元前 3000 年ごろ	紀元前 2100 年ごろ	紀元前 700 年ごろ
lu <sub>2</sub> 「人」 人の形をかたどる。			
gu <sub>2</sub> 「食べる」 ka 「口」と ninda 「パン」の文字であらわされる。			
maš 「半分」			
ba <sub>2</sub> 「城壁」 外側が城壁をかたどり、中の文字が音符をあらわす。			
sum 「タマネギ、与える」 本来、タマネギをあらわす文字 sum が同音の「与える」も意味する。			
apin 「すき、農夫」 本来、apin 「すき」をあらわす文字が engar 「農夫」としても用いられる。			

森若葉「楔形文字で日本語を書く」『月刊みんぱく』2005年10月号, p.17

## 楔形文字って何？①

- メソポタミアのシュメール人が発明した最古の文字
- 紀元前4千年紀末～紀元後2/3世紀の中近東世界でさまざまな言語を表記するために使用された(西は地中海～東はペルシア湾、北はアナトリア～南はエジプト)
- 何が記録されたのか？  
王碑文・奉納碑文・行政経済文書・契約書・裁判記録・書簡・条約・占い・呪術・数学・天文学・文学作品・ことわざ・語彙リスト・地図...



## 楔形文字って何？②

- 楔形文字  
楔のような形の画を組み合わせた文字  
縦・横・斜めの楔と、楔の頭で構成される
- 筆記方法  
主に葦の尖筆を使って粘土の上に刻まれた
- ✓表語(表意)文字も表音文字もある
- ✓時代により文字も変化(象形文字→簡略化)



### 楔形文字で記された言語 ～シュメール語～

lugal-e      王は  
dingir-ra    神のために  
lagaš-a      ラガシュに  
e<sub>2</sub>            神殿を  
mu-na-ni-in-du<sub>3</sub>    建てた。

### 楔形文字で記された言語 ～ヒッタイト語～

nu NINDA-an    e-ez-za-at-te-ni    wa-a-tar-ma    e-ku-ut-te-ni  
「あなたたちはパンを食べ、水を飲む」



## 今後の研究 古代トルコの王国ヒッタイトの領域と境界について

### ➤ヒッタイト王国: 前2千年紀アナトリアの王国



### 研究方法

- ①「支配」や「境界」を意味するヒッタイト語の用例研究  
→どこからどこまで支配していると考えていたのか
- ②国境付近のモニュメントと碑文の調査  
→実際の境界はどこにあったのか？

### 関連する語の用例研究

- ✓manniyahh-(動詞)  
「支配する」/「支配を委ねる」
- ✓ZAG (ヒッタイト語でirḫa-)(名詞)  
「境界」



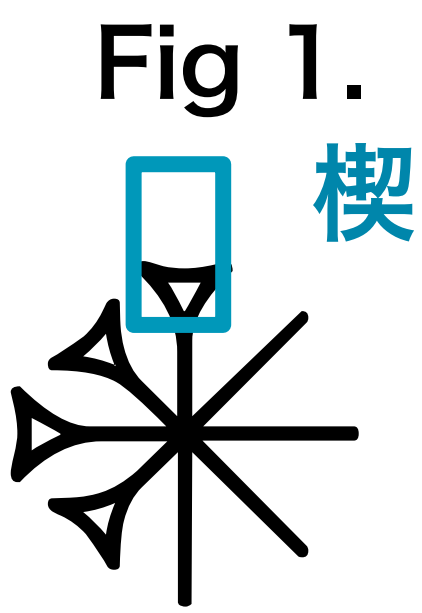


# 楔形文字文献の形態論情報付き コーパス構築の自動化に向けて



## 背景

- ❖ 楔形文字
  - ✓ 中近東で紀元前4千年紀から約**3,000年間**利用された文字
  - ✓ 約600種類 (時代により字形が異なる)
  - ✓ 表音文字にも表意文字にも使用
  - ✓ 外見上「**楔**」と線からなる (Fig 1.)
  - ✓ 主に粘土板に書かれた (Fig 2.)
- ❖ 楔形文字文書は言語・文献学的に重要
  - ✓ 古代中近東の言語資料性
  - ✓ 文献ジャンルの多様性
- ❖ 最低でも約**50万件**の文献が存在



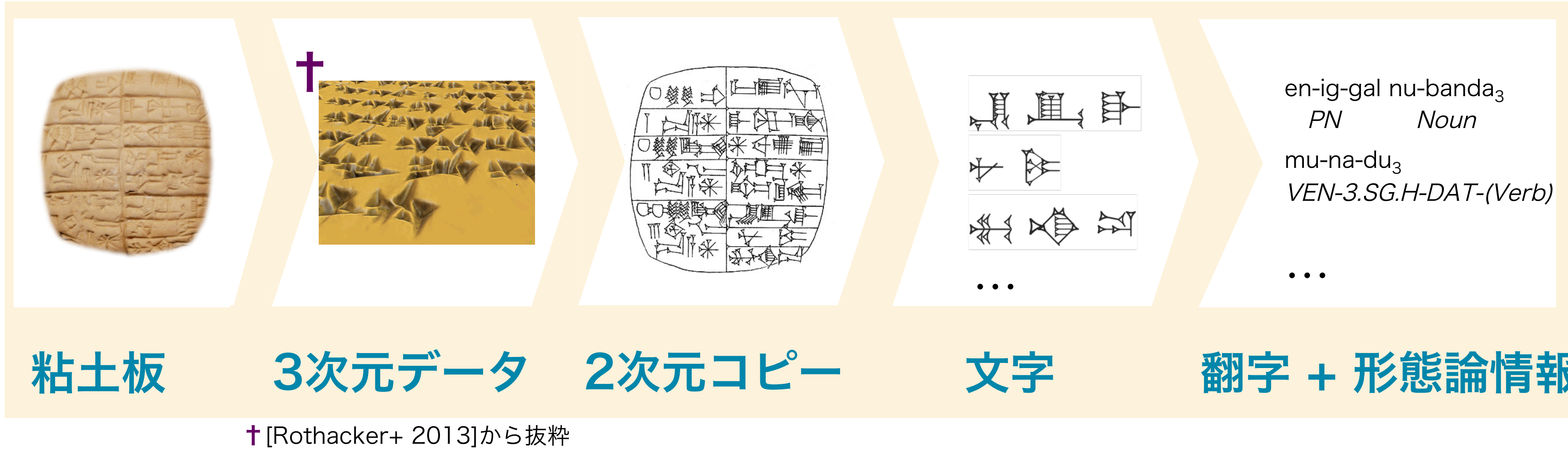
- ❖ 効率的な文献調査には電子コーパス化が不可欠
- ❖ 文法(形態論)情報のあるコーパスの構築は進んでいない
- ❖ **手作業でのコーパス構築の労力を省くため、計算機を利用した自動構築を目指す**

## 関連研究

現在形態論情報付きコーパスの自動構築を扱った研究はほとんどない

- ❖ 粘土板の3次元データ化と2次元コピー画像の作成  
[Watkins+ 2003], [Mara+ 2013], [Rothacker+ 2013] etc.
- ❖ 手作業での形態論情報付きコーパス構築  
[Molina+ 2016], ETCsRI (<http://oracc.museum.upenn.edu/etcsri>) etc.

## 形態論情報付きコーパス自動構築の 想定フロー



- ❖ 現時点では文字の認識すら実現していない
- ❖ 研究利用できる2次元コピー画像は多く存在する
- ➔ **まず2次元コピー画像からの文字種認識を目指す**

## 楔形文字画像データセットの構築

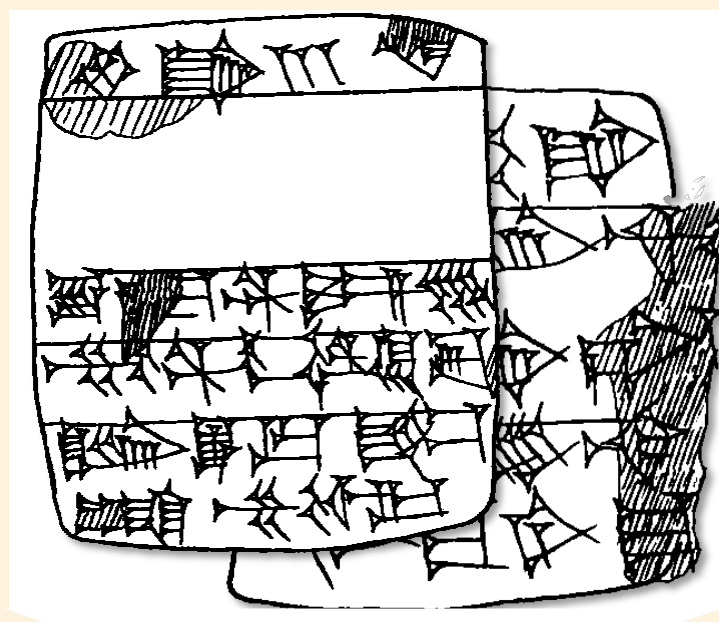
- ❖ 文字種認識実現には、文字ごとに一定規模の画像データが必要  
➔ 2次元コピー画像から半自動で収集
- ❖ 高頻度で出現する100文字  
それぞれに200画像の収集を目標
- ❖ データセット構築フロー
  1. 2次元コピー画像群からヒューリスティックに文字(領域)候補群を抽出
  2. 文字候補画像群をt-SNEで次元圧縮し、DBSCANでクラスタリング
  3. 手作業でクリーニング

- ❖ これまでの収集結果は公開中

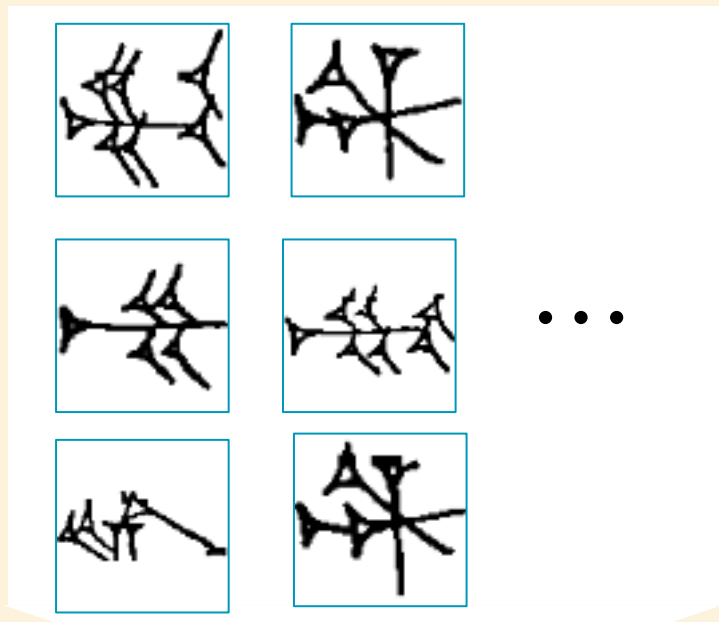
## 今後

- ❖ 手書きコピー画像からの文字認識
- ❖ シュメール語で、文字の並びからの形態論情報付与

### 2次元コピー



### 文字候補群



### 分類結果

